SUMMARY AND GROUP CORRELATION

The possibility of finding the arithmetic mean of several groups separately, and combining the data in order to find the arithmetic mean of the combined groups, is well known and is frequently used to save the labor of additional computations. It may not be generally known that an analagous process is available in the case of the coefficient of correlation.

Let x and y be the deviations, and A and B the averages of two series, the original values being expressed by X and Y.

Then
$$\begin{aligned}\Sigma xy &= \Sigma(X-A)(Y-B)\\ &= \Sigma XY + nAB - A\Sigma Y - B\Sigma X\\ &= \Sigma XY + \frac{n(\Sigma X)(\Sigma Y)}{n\ \ n} - \frac{(\Sigma X)(\Sigma Y)}{n} - \frac{(\Sigma X)(\Sigma Y)}{n}\\ &= \Sigma XY - \frac{(\Sigma X)(\Sigma Y)}{n}\end{aligned}$$
$$\begin{aligned}\Sigma x^2 &= \Sigma(X-A)^2\\ &= \Sigma X^2 - 2A\Sigma X + nA^2\\ &= \Sigma X^2 - 2\frac{(\Sigma X)(\Sigma X)}{n} + n\frac{(\Sigma X)^2}{n^2}\\ &= \Sigma X^2 - \frac{(\Sigma X)^2}{n}\end{aligned}$$

and, similarly, $\Sigma y^2 = \Sigma Y^2 - \dfrac{(\Sigma Y^2)}{n}$.

Hence
$$\begin{aligned}r &= \frac{\Sigma xy}{n\sigma_x\sigma_y} = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}}\\ &= \frac{\Sigma XY - \dfrac{(\Sigma X)(\Sigma Y)}{n}}{\sqrt{\left[\Sigma X^2 - \dfrac{(\Sigma X)^2}{n}\right]\left[\Sigma Y^2 - \dfrac{(\Sigma Y)^2}{n}\right]}}\\ &= \frac{n\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{[n\Sigma X^2 - (\Sigma X)^2][n\Sigma Y^2 - (\Sigma Y)^2]}}.\end{aligned}$$

Since X and Y may be any scalar values, the items may be grouped into classes, and arbitrary values may be assigned to the classes, *e. g.*, $-2, -1, 0, +1, +2. \ . \ . \ .$

If in two or more groups the same arbitrary class values are used, the number of items in one group being n_1, in another n_2, etc., the correlation coefficient for each group may be obtained separately, and by a simple computation the correlation coefficient for the combined groups may be obtained:

$$r = \frac{(n_1+n_2+\ .\ .\ .)\Sigma(\Sigma XY) - [(\Sigma X)(\Sigma Y)]}{\sqrt{\{(n_1+n_2+\ .\ .\ .)\Sigma(\Sigma X^2) - \Sigma[(\Sigma X)^2]\}\{(n_1+n_2+\ .\ .\ .)\Sigma(\Sigma Y_2) - \Sigma[(\Sigma Y)^2]\}}}.$$

Herbert A. Sturges

University of Washington